KWANZA TUKULE DATA ANALYST ASSESSMENT REPORT

SUBMITTED TO :

KWANZA TUKULE

Contents

Data Cleaning and Preparation

Executive Summary

This report provides a detailed analysis of the sales and customer data, focusing on identifying key trends, anomalies, customer segmentation, and strategic insights. Key performance indicators (KPIs) include total sales value, total quantity sold, sales growth rate, and correlation between quantity and sales value. The analysis concludes with actionable recommendations for product prioritization, customer retention, and operational efficiency.

1.1 Data Quality Assessment

The dataset was assessed for quality, and the following issues were identified:

- Missing Values: Some records contained missing data in the "Sales Value" columns. These gaps were filled by imputation using the Median value for each respective product.
- Duplicates: Duplicate records were found and the duplicates were removed to ensure the dataset's integrity.
- Inconsistent Data Types: The "DATE" column had entries stored as text, which were reformatted to a datetime type for accurate analysis.

1.2 Feature Engineering

A new column, "Month-Year," was created by extracting the month and year from the "DATE" column. The formula used in Excel was:
=TEXT(DATE, "mmm-yyyy")
This new column allows for grouping data by time period and facilitates trend analysis.

| DATE	MONTH-YEAR	ANONYMIZED CATEGORY	ANONYMIZED PRODUCT	ANONYMIZED BUSINESS	ANONYMIZED LOCATION	QUANTITY	UNIT PRICE		ACTUAL TOTAL SALES
1/18/2024	January 2024	Category-77	Product-d09c	Business-783b	Location-128a	1.00	KES 3,180	KES	3,180
1/2/2024	January 2024	Category-120	Product-4156	Business-fef1	Location-3e32	2.00	KES 2,180	KES	4,360
1/2/2024	January 2024	Category-85	Product-d06e	Business-fef1	Location-3e32	2.00	KES 1,130	KES	2,260
1/23/2024	January 2024	Category-83	Product-0367	Business-25ce	Location-1979	1.00	KES 3,600	KES	3,600
1/23/2024	January 2024	Category-75	Product-2175	Business-25ce	Location-1979	1.00	KES 4,460	KES	4,460
1/23/2024	January 2024	Category-120	Product-9a3e	Business-25ce	Location-1979	1.00	KES 2,130	KES	2,130
1/7/2024	January 2024	Category-102	Product-4883	Business-77af	Location-b27b	2.00	KES 230	KES	460
1/7/2024	January 2024	Category-76	Product-e805	Business-77af	Location-b27b	1.00	KES 7,790	KES	7,790
1/1/2024	January 2024	Category-120	Product-29ee	Business-22e2	Location-b125	20.00	KES 2,120	KES	42,400
1/1/2024	January 2024	Category-120	Product-7493	Business-22e2	Location-b125	5.00	KES 2,140	KES	10,700
1/1/2024	January 2024	Category-120	Product-dfdf	Business-22e2	Location-b125	5.00	KES 2,120	KES	10,600
1/12/2024	January 2024	Category-85	Product-0c64	Business-e0ec	Location-e2f8	1.00	KES 1,190	KES	1,190
1/12/2024	January 2024	Category-75	Product-2175	Business-e0ec	Location-e2f8	2.00	KES 4,410	KES	8,820
1/12/2024	January 2024	Category-119	Product-66e0	Business-e0ec	Location-e2f8	2.00	KES 1,740	KES	3,480
1/12/2024	January 2024	Category-77	Product-d09c	Business-e0ec	Location-e2f8	1.00	KES 3,155	KES	3,155
1/12/2024	January 2024	Category-76	Product-e805	Business-e0ec	Location-e2f8	1.00	KES 7,500	KES	7,500
1/1/2024	January 2024	Category-100	Product-94a8	Business-b299	Location-d1a8	2.00	KES 2,200	KES	4,400
1/1/2024	January 2024	Category-77	Product-a021	Business-b299	Location-d1a8	2.00	KES 2,870	KES	5,740

2. Exploratory Data Analysis

2.1 Sales Overview

The total sales were analyzed across different dimensions:

Total Quantity and Value by Anonymized Category:

The data was grouped by "Anonymized Category," and total sales quantity and value were calculated for each category. A bar chart was created to compare total Quantity and Value by category.
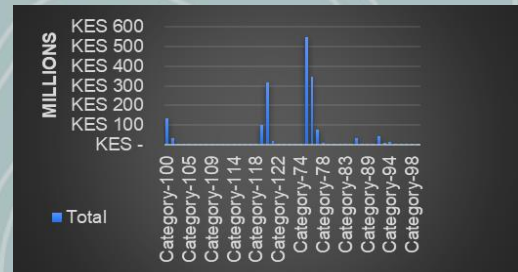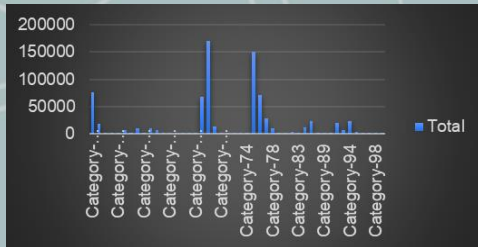
Total Quantity and Value by Anonymized Business:

The data was grouped by "Anonymized Business" to calculate total sales value and quantity. A table was created to highlight the top-performing businesses in terms of both value and quantity.
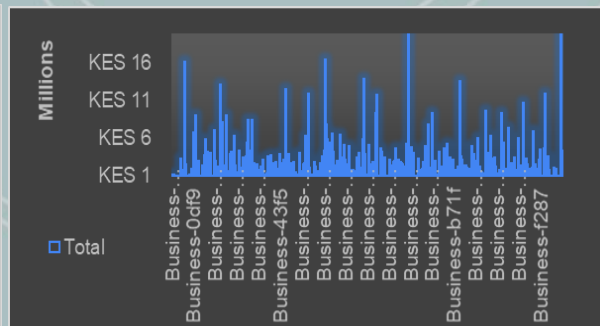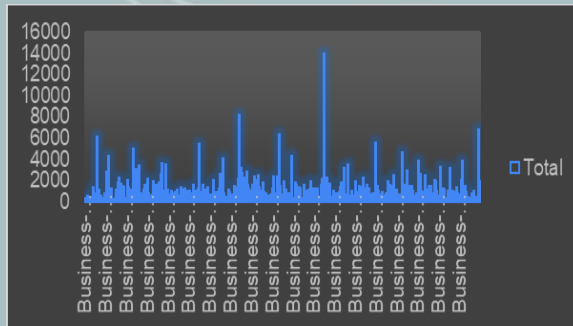
2.2 Trends Over Time

A time series analysis was conducted to observe the seasonal trends of sales:

Monthly Trends in Quantity and Value: A time series plot was generated for "Sales Value" and "Quantity" over each "Month-Year." The analysis showed significant fluctuations, with peaks in April and September, and a notable decline in December 2024.
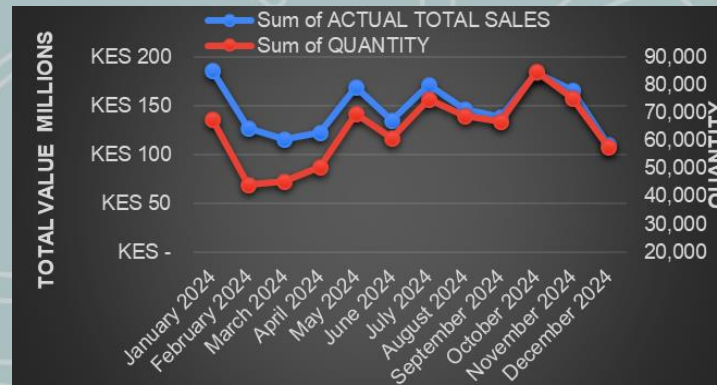
2.3 Performance Analysis

Top 5 Products by Quantity: The products with the highest quantities sold were identified, with Product-e805 being the leader.

Top 5 Products by Sales Value: The top 5 products based on sales value were determined, with Product-e805 again leading both in quantity and value.

| Anonymized Product	TOTAL QUANTITY	PERCENTAGE QUANTITY
Product-29ee	35940	18.76%
Product-4156	28487	14.87%
Product-66e0	46957	24.51%
Product-8f75	37566	19.61%
Product-e805	42602	22.24%
Grand Total	191552	100.00%

Anonymized Product	TOTAL SALES	PERCENTAGE
Product-29ee	KES 68,248,274	11.05%
Product-4156	KES 56,956,007	9.22%
Product-66e0	KES 70,704,225	11.45%
Product-8f75	KES 158,797,460	25.72%
Product-e805	KES 262,787,281	42.56%
Grand Total	KES 617,493,247	100.00%

Performance Visualization: A bar chart was used to display the top 5 products by both quantity and value, showing clear leaders in both categories.
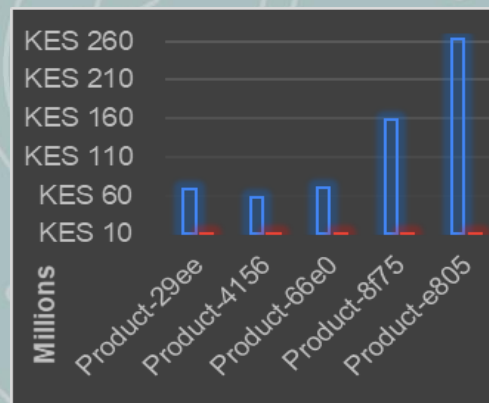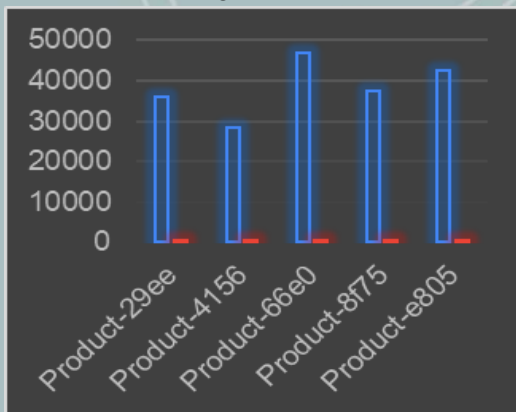
Insights

1. **Product-e805** leads both in total quantity (22.24%) and sales value (42.56%), making it the most valuable product for the business.
2. **Product-8f75** ranks second in terms of sales value (25.72%) but fourth in quantity sold, indicating it is a high-value product despite lower sales volume.
3. **Product-66e0** has a balanced contribution to both total quantity (24.51%) and sales value (11.45%), suggesting it has consistent demand across the market.

Strategic Recommendation

- **Focus Campaigns on Product-e805:** Amplify marketing campaigns around Product-e805, leveraging its strong performance in both metrics.
- **Evaluate Pricing and Promotion for Product-8f75:** Despite contributing significantly to sales value, it has lower sales volume compared to other products. This indicates potential for further growth with targeted offers or discounts.
- **Monitor Inventory for Product-66e0:** Its consistent demand highlights the need for adequate stock levels to meet customer requirements.

3. Advanced Analysis

3.1 Customer Segmentation

Businesses were segmented based on their purchasing behavior:

Segmentation by Total Quantity Purchased:

- High-Quantity: 34.1%
- Medium-Quantity: 31.1%
- Low-Quantity: 34.8%

Total Quantity

1491 1637

1672
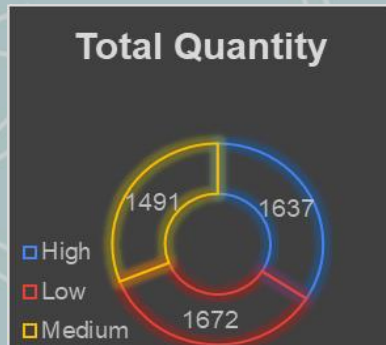
☐ High
☐ Low
☐ Medium

Segmentation by Transaction Frequency:

- High-Frequency: 33.5%
- Medium-Frequency: 30.9%
- Low-Frequency: 35.7%

Frequency of Transactions

1481 1606

1713
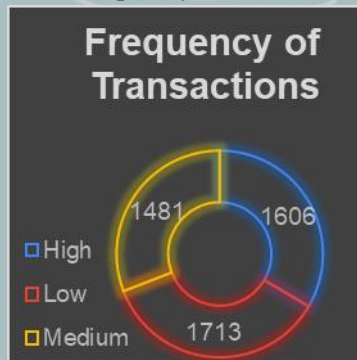
☐ High
☐ Low
☐ Medium

Segmentation by Total Sales Value:

- **High-Value: 34.3%**
- Medium-Value: 32%
- Low-Value: 33.8%

Total Value

1535 1645

1620
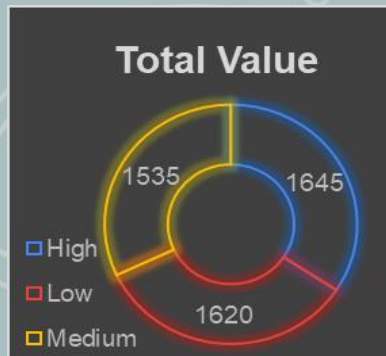
☐ High
☐ Low
☐ Medium

- **Balanced Distribution**: There is an even distribution across High, Medium, and Low classifications in terms of quantity, frequency, and value, suggesting opportunities for targeted strategies in each segment.
- **High-Value and High-Frequency Customers**: These customers represent the most profitable segment. Special engagement strategies, like loyalty programs or exclusive offers, could strengthen their brand loyalty.
- **Low-Frequency Customers**: This group has the potential for improvement. Regular campaigns or reminders to purchase could re-engage this segment.

Strategic Recommendations

- **High-Value Customers**:
  - Develop VIP programs to reward their loyalty and encourage repeat transactions.
  - Offer tailored promotions to upsell or cross-sell products.
- **Medium-Value Customers**:
  - Educate this group about premium products or bundles to boost their value contribution.
  - Introduce referral incentives to help increase their engagement.
- **Low-Value Customers**:
  - Provide entry-level incentives, such as discounts or promotions, to encourage higher spending.
  - Send personalized communication to understand their needs better and improve their buying patterns.

3.2 Forecasting

A 3-month sales forecast was generated using the Exponential smoothing method. The forecasted sales for January, February, and March 2025 show a gradual recovery from the decline in December 2024:

- January 2025: KES 124,198,336
- February 2025: KES 124,255,308
- March 2025: KES 124,312,280
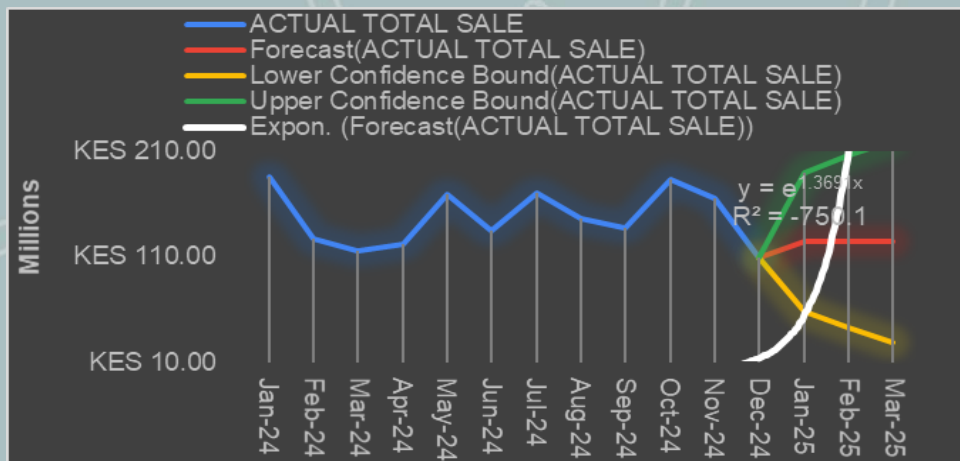
Insights:
- Sales forecasts for **January** to **March 2025** indicate a potential recovery from the December 2024 decline.

- There is a large confidence interval in the forecast values, particularly with the lower bounds, indicating uncertainty in these months.
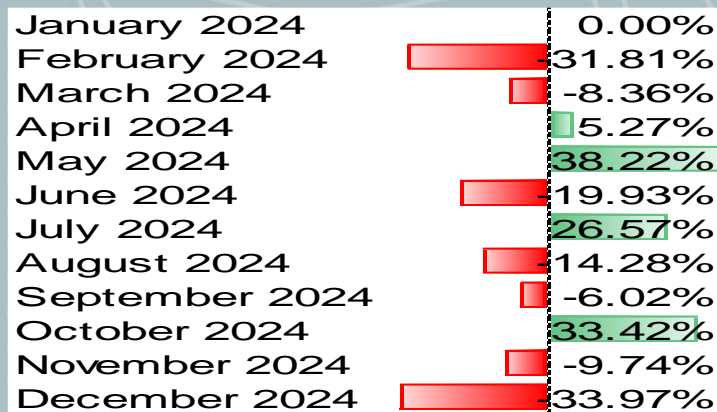
- **Align Sales Strategy to Forecasts**: Prepare for varying levels of sales with flexible inventory and marketing plans to handle potential fluctuations.
- **Promotions in Low Sales Months**: Implement targeted campaigns in traditionally lower sales months (e.g., **February** and **December**) to mitigate dips.

3.3 Anomaly Detection

Anomalies in sales were identified based on significant deviations:

December 2024: A sharp decline in sales (-33.97%) was identified, likely due to external factors such as holiday seasonality or supply chain disruptions.

May and October 2024: Significant spikes in sales were observed, likely driven by seasonal demand.

Month	Change
January 2024	0.00%
February 2024	-31.81%
March 2024	-8.36%
April 2024	5.27%
May 2024	38.22%
June 2024	-19.93%
July 2024	26.57%
August 2024	-14.28%
September 2024	-6.02%
October 2024	33.42%
November 2024	-9.74%
December 2024	-33.97%

3.4 Correlation Analysis

The correlation between "Quantity" and "Actual Total Sales Value" was found to be 0.835, indicating a strong positive relationship. This suggests that strategies aimed at increasing quantity sold would likely result in higher sales values.

4. Strategic Insights and Recommendations

Product Strategy
- **Prioritize Product-e805**: Invest in marketing and inventory for Product-e805 to capitalize on its popularity.
- **Diversify Product Portfolio**: Enhance visibility and marketing of mid-performing categories like Product-b456 to spread revenue across more categories.

Customer Retention
- **Re-Engage Medium-Value Customers**: Offer loyalty programs or discounts to medium-value customers to encourage repeat purchases.
- **Target Low-Value Customers**: Provide personalized campaigns or bundle offers to increase their purchase volume.

Operational Efficiency

- **Inventory Management**: Ensure adequate stock levels of high-performing products, especially during peak seasons like October.
- **Forecasting Tools**: Use advanced forecasting methods to predict demand and minimize overstock or understock issues.

Sales Growth

- **Upselling Strategies**: Train the sales team to suggest complementary products to increase the average transaction value.
- **Seasonal Promotions**: Launch promotions in months with historically low sales, like December, to boost performance.

5. Dashboard and Reporting

An Excel dashboard was created to summarize key insights, including:

- Total Quantity and Value by Anonymized Category
- Top-performing Products and Businesses
- Time-series chart of Sales Trends
- Segmentation Summary of Customer Groups

The dashboard was created using a combination of bar charts, tables, and time-series plots, allowing for interactive data exploration within Excel.

5.1 Key Features of the Dashboard:

Sales Overview: Displays total sales values and quantities for each business category.
Performance Summary: Shows the top products by quantity and sales value.
Customer Segmentation: Visualizes customer groups based on purchase behavior.
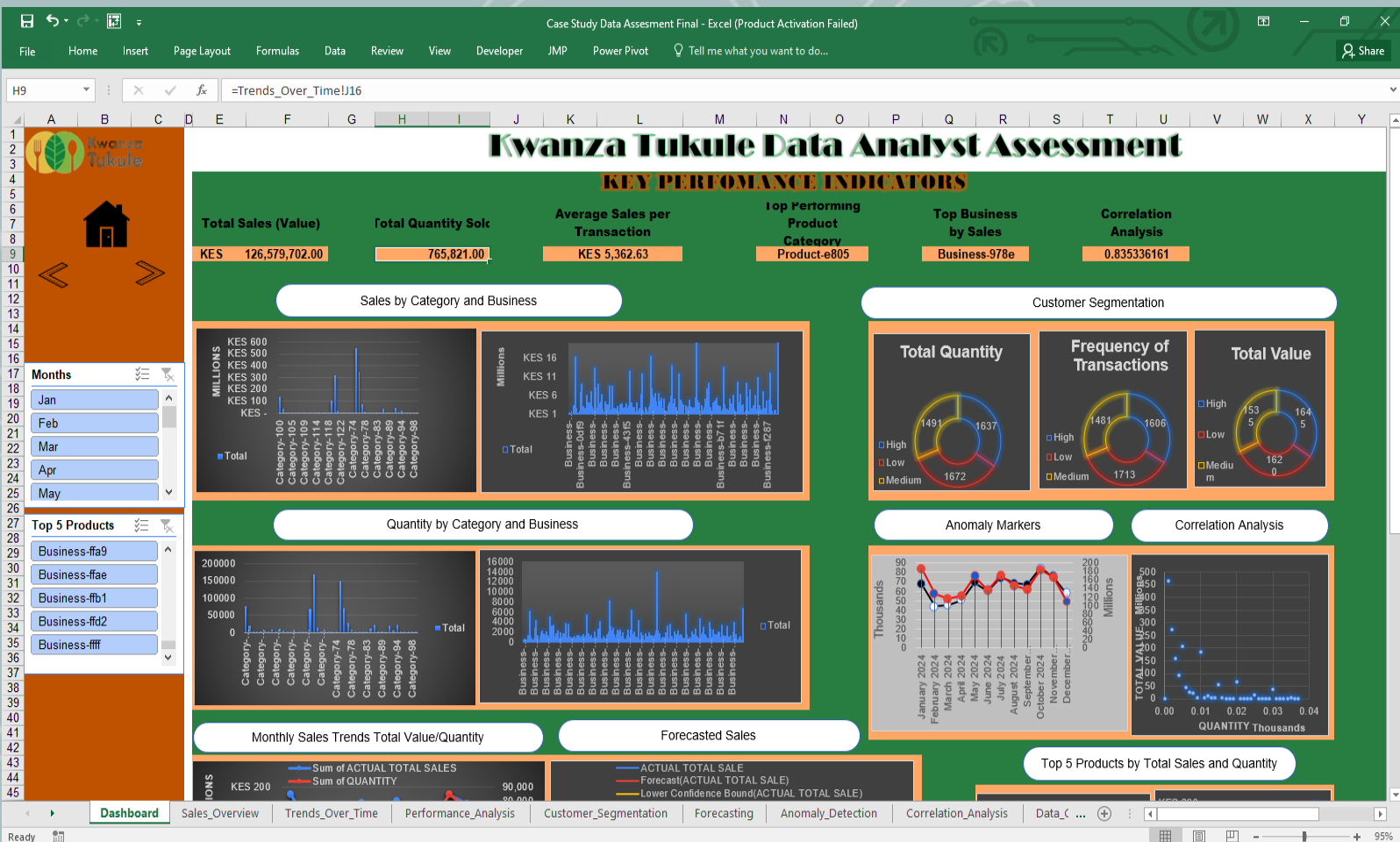
5.2 Scalability

If the dataset were 10 times larger, the following optimizations would be recommended:

- **Data Storage**: Use cloud storage (e.g., Amazon S3) for large datasets to ensure efficient access and scalability.
- **Data Processing**: Utilize distributed processing frameworks like Apache Spark to handle large-scale data analysis.
- **Analysis**: Optimize Excel processing by segmenting data into manageable chunks or exploring more advanced tools like Python or R for large dataset handling.